AD_____
(Leave blank)

Award Number:

W81XWH-10-2-0128

TITLE:

ENGINEERING ENVIRONMENTALLY-STABLE PROTEASES TO SPECIFICALLY NEUTRALIZE PROTEIN TOXINS

PRINCIPAL INVESTIGATOR:

Philip N. Bryan

CONTRACTING ORGANIZATION:

Potomac Affinity Proteins, LLC
NORTH POTOMAC MD 20878-2566

REPORT DATE:

October 2013

TYPE OF REPORT:

Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT:

X Approved for public release; distribution unlimited

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. REPORT DATE *(DD-MM-YYYY)* | 2. REPORT TYPE | 3. DATES COVERED *(From - To)* |
|---|---|---|
| October 2013 | Final | 15September2010-14September2013 |

**4. TITLE AND SUBTITLE**

ENGINEERING ENVIRONMENTALLY-STABLE PROTEASES TO SPECIFICALLY NEUTRALIZE PROTEIN TOXINS

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**
W81XWH-10-2-0128

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Philip N. Bryan

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Potomac Affinity Proteins, LLC
11305 DUNLEITH PL
NORTH POTOMAC MD 20878-2566

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

U.S. Army Medical Research Materiel Command
Fort Detrick, MD 21702-5012

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION / AVAILABILITY STATEMENT**

Approved for public release, distribution unlimited

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

This project is intended to develop the tools and principles necessary to engineer subtilisin proteases which specifically target and deactivate biological warfare agent (BWA) toxins. We are engineering and evolving subtilisin proteases that specifically target and deactivate BoNT, SEB, ricin, and B. anthracis lethal factor (LF), representing four functionally distinct families of toxins. The centerpiece of our design effort is a phage-display selection method for creating tightly-regulated proteases of high specificity. In this system the protease, substrate sequence, and regulatory co-factor are co-evolved. The key accomplishments this past year were: 1. Determined the structure of an evolved variant pT2077 in complex with the substrate sequence used to select it. 2. Design/evolution of a highly active enzyme that can cut P4 = I (pT2050). 3. Computational design of specificity for an ionic P4 amino acid (P4 = E, pT2121 and P4 = K, pT2114); 4. Engineered protease chain reactions that can reliably measure concentrations of 250 fM range in a 20 hour assay.

**15. SUBJECT TERMS**

Enterotoxin, protease, directed evolution, subtilisin, protein engineering, phage-display, enzymology

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| | | | | | USAMRMC |
| a. REPORT | b. ABSTRACT | c. THIS PAGE | UU | 17 | 19b. TELEPHONE NUMBER *(include area code)* |
| U | U | U | | | |

## Table of Contents

## INTRODUCTION

This project was intended to develop the tools and principles necessary to engineer subtilisin proteases that specifically target and deactivate biological warfare agent (BWA) toxins. We have engineered and evolved subtilisin proteases to specifically target and deactivate BoNT, SEB, ricin, and B. anthracis lethal factor (LF), representing four functionally distinct families of toxins.

Developing principles for engineering enzymatic function will lead eventually to enormously powerful, biologically-inspired materials. Serine proteases are among the most studied and best-understood enzymes and offer unique opportunities for progress. Serine proteases of the chymotrypsin and subtilisin families became early model systems for protein engineering because of well-characterized mechanisms, timely cloning of the genes, ease of expression and purification, and the availability of high-resolution atomic resolution structures. There are several excellent reviews of these early studies (*1-3*). Although the concept of evolving protease specificity might appear simple, the mechanistic knowledge of proteases required to engineer their specificity turns out to be very complex. Substrate-enzyme interactions are well characterized for subtilisin from high resolution x-ray structures of many protease-inhibitor complexes (*4-7*). At first glance, engineering protease specificity may seem to be a problem of engineering lock and key fit between the protease and the substrate sequence one desires to cut. We observe, however, sequence-specific cleavage is much more subtle, depending upon how side chain interactions influence not only ground state binding but also the positioning in the scissile bond relative to catalytic amino acids.

In subtilisin, most contacts are with the first five substrate amino acids on the acyl side of the scissile bond (denoted P1 through P5, numbering from the scissile bond toward the N-terminus of the substrate (*8*)) and the first amino acid on the leaving group side (denoted P1'). The backbone of the substrate inserts between strands 100-104 and 125-129 of subtilisin to become the central strand in an anti-parallel b-sheet arrangement involving ten main chain H-bonds (*9, 10*). Hence, a major component of substrate binding energy involves the peptide backbone. The side chain components of substrate binding result primarily from the P1 and P4 amino acids (*11-13*). Optimal substrates for subtilisin have large hydrophobic amino acids at the S1 and S4 sub-sites of the enzyme (*11, 12*).

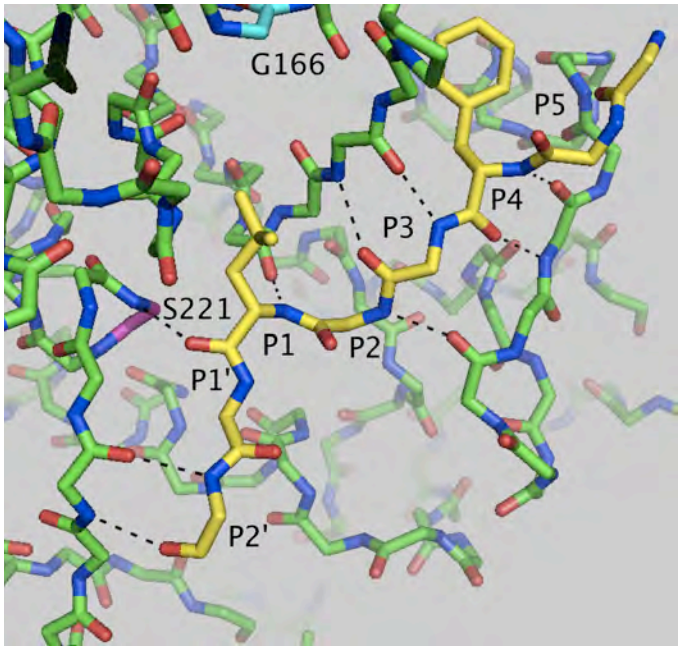


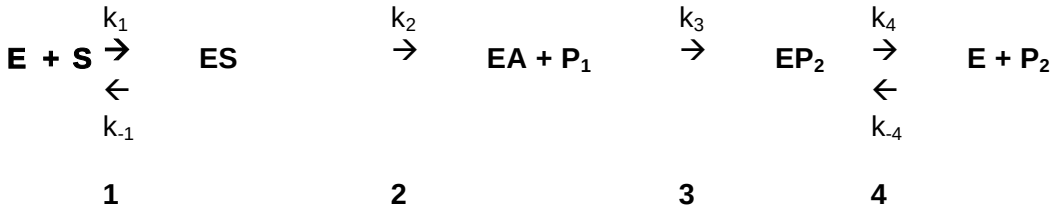

***Figure 1.*** *Structure of a peptide substrate (yellow) spanning the subtilisin active site. Black dashed lines represent interactions represent main chain H-bonds between the peptide and the subtilisin binding cleft. The side chains of the P1 leucine and the P4 phenyalanine and shown. The position of the catalytic serine 221 is shown in pink as well as glycine 166 at the back of the S1 pocket. The depiction is based on 3BGO.pdb* (*14*).

In order to engineer toxin-specific proteases, we identified target amino acid sequences in protein toxins and then engineered high-specificity proteases against the selected sequences. The design/selection effort had five elements: 1) identify cognate sequences from target toxins that can be cut with prototype subtilisins (USAMRIID); 2) create specificity for cognate sequences by design/evolution; 3) confirm proteolysis on intact toxins (USAMRIID), and 4) test catalytic properties of new proteases; 5) test the ability of engineered proteases to deactivate the selected toxins in vivo (USAMRIID).

## BODY

### *Evolving tunable chemistry*

Below is a minimal realistic mechanism for peptide hydrolysis by a serine protease:




The reaction can be divided into four phases: 1) substrate binding; 2) acylation and release of the C-terminal peptide ($P_1$), 3) deacylation and 4) dissociation of the N-terminal peptide ($P_2$). Nucleophilic attack of the carbonyl carbon of the scissile amide bond is carried out by the active site serine. The other two amino acids forming the catalytic triad are histidine and aspartic acid which form a charge relay system. Serine proteases have evolved to manage the burial of charged groups during the catalytic cycle. In the enzyme-substrate complex, the catalytic aspartic acid forms a very strong H-bond to Nd1 of histidine which polarizes the histidine and allows Ne2 to act as a proton shuttle during acylation and deacylation reactions. Our approach to evolving high specificity proteases rests on the premise that the active site aspartic acid (D32) can be mutated such that exogenous anions can rescue activity and that anion concentration can control the flux of substrates, transition states, intermediates and products through the reaction pathway to maximize sequence specificity.

Typically steady state kinetic measurements are used to assess the specificity of a protease. Specificity is usually defined as the ratio of $k_{cat}/K_M$ of an enzyme for one substrate relative to another. Determining $k_{cat}/K_M$ values for two substrates allows quantitation of sequence preferences but does not reveal the kinetic and thermodynamic basis for the preference (*15*). To understand the mechanistic basis for specificity, transient state kinetic methods must be employed to determine microscopic rate constants. It is important to understand that $K_M$ and $k_{cat}$ are composite rate constants into which are folded multiple microscopic rate constants for the multi-step hydrolysis reaction. It frequently is assumed for many enzymatic reactions that $k_{cat} \sim k_2$ and $K_M \sim K_S$. These relationships are accurate only if $k_2$ is small compared to $k_{-1}$, $k_3$ and $k_4$ however. As $k_2$ approaches $k_{-1}$, substrate binding can no longer be viewed as a rapid equilibrium which is kinetically uncoupled from acylation. This has important consequences for specificity. The $k_{cat}/K_M$ value is the apparent second order rate constant for productive substrate binding. It is less than the true binding rate ($k_1$) by a factor of $k_2/(k_{-1} + k_2)$ (*15*). As $k_{-1}$ slows to less than the acylation rate and the enzyme begins to reach a maximum determined by the rate of substrate binding, as the coefficient $k_2/(k_{-1} + k_2)$ approaches one. Thus coupling between substrate binding and acylation (the first chemical step) broadens specificity. Further, as product release becomes slower than acylation, it determines the $k_{cat}$ of the reaction rather than the acylation rate.

Examination of the microscopic rate constants for anion-triggered cleavage reactions shows how controlling the flux of species through the pathway favors sequence discrimination. There are three important observations: 1) The $k_{cat}/K_M$ for the optimal substrate is significantly less than the substrate binding rate; 2) The substrate dissociate rate is faster than the acylation rate ; 3) The deacylation and product dissociation rates are > than the acylation rate (i.e. no burst kinetics are observed). As a result, the specificity is influenced by its affinity for the different substrates, as well as the effect of substrate sequence on the acylation rates. The general conclusion is that tuning the chemistry to match binding steps is necessary to achieve optimum specificity.

**Phage-display selection methods for creating anion-regulated proteases of high specificity**

The first step in the directed evolution of high specificity proteases is identifying a regulatory anion that can control subtilisin activity in the selection process. In order to achieve efficient hydrolysis, the scissile bond of the substrate, the catalytic residues of the enzyme (H64, N155 and S221) and the anion must be brought into precise register. Co-evolving the anion site with a specific substrate sequence optimizes this positioning and leads to more efficient turn-over of the co-evolved substrate. S189 was the starting point for evolving new specificities. We decided to focus on azide and nitrite anions for two reasons. They are small enough fit into the space in the active site created by mutation of D32 and their pKa's are high enough to allow adequate binding to the ground state and low enough to provide strong polarization of H64.

*1) Design of a refined random library for anion triggering.* The theory of random library design is that a proper constellation of neighboring residues can create selective binding pockets for substrate amino acids and specific anions. The amino acids chosen for randomization in the anion site library are 30, 32, 33, 62, 68 and 125 (Figure 2).

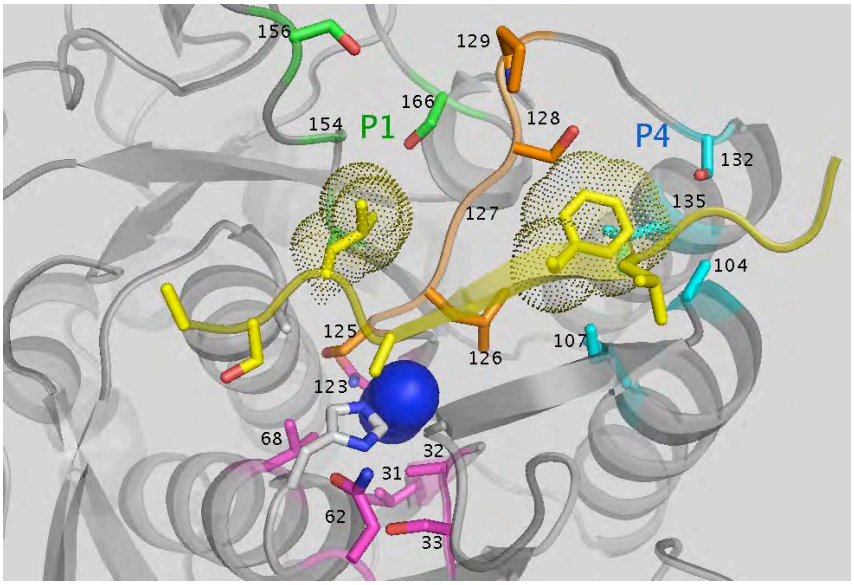



**Figure 2.** Sites of random mutagenesis in the P1 pocket are in green, sites in the P4 pocket in cyan, and sites in the anion pocket are in violet. P1 leucine and P4 phenylalanine are shown with dot surfaces. The three binding sites are interconnected by common amino acids in the region form 123-129. These amino acids are in orange.

*2) Phagemid vector development* Vector development for phage display involved three modifications to existing phagemid vectors: 1) introducing a pTac promoter into pHen; 2) employing an amber codon at Q10 of mature subtilisin instead of between subtilisin and G3P; 3) using a refined strategy for transfection and growth which improve genetic stability of fusion proteins. pHen vectors developed for this project are shown in *Appendix 1: pHen vectors, page 18*.

*3) Using a catch and release phage display method to evolve a binding site for nitrite which will trigger the cleavage of a cognate sequence.* The "catch" phase of phage selection was carried out using a fusion protein comprising an albumin-binding domain ($G_A$), an engineered subtilisin prodomain containing the cognate sequence ($P_{LFRAL-S}$) (*16*), and an IgG–binding domain ($G_B$). In this scheme the subtilisin was synthesized as a fusion protein on the surface of M13 phage. The random library of subtilisin phage is mixed with the $G_A$- $P_{LFRAL-S}$-$G_B$ substrate. Phage which display a misfolded subtilisin or one which has subsites which bind poorly to the target sequence are rejected on the basis of non-binding. Phage which bind to substrate are in turn bound to IgG sepharose via the $G_B$ domain in the catch step. Subtilisin phage which cleave the substrate without the trigger are not retained in the catch step of the selection. In subtilisin phage which perform the acylation step in response to the nitrite trigger, the ternary complex is converted into an acyl-enzyme with the concomitant release of the $G_B$. The rate of release of a particular subtilisin-phage reflects both its affinity for anion and the ability of the anion to stabilize the transition state for acylation. Thus we are able to select the two major energetic components contributing to specificity. The phage released from IgG Sepharose (but still tightly bound to $G_A$-$P_{COGNATE}$) are then collected on HSA Sepharose. Finally, the subtilisin phage which both bind and cleave the cognate sequence are eluted from the HSA Sepharose at pH 2.5. The theory of random library design is that a proper constellation of neighboring residues can create selective binding pockets for substrate amino acids and a triggering anion.

Results are reported in detail in Annual reports for 2011 and 2012. Selected mutants subjected to kinetic analysis are shown in *Appendix 2: Primary anion screening, Appendix 3: Secondary anion screening with variable P2, and Appendix 4: Tertiary anion screening with variable P1, pages 19-22.*

Substrates used in analysis are shown in *Appendix 5: P2 substrate series and Appendix 6: P1 and P4 substrate series, pages 23-25.*

*4) Structure determination of a refined anion-triggered variant* We have determined the high-resolution x-ray structure of the evolved variant pT2077 in complex with the cognate peptide LFRAL.

**Phage-display selection methods for evolving subtilisin sub-sites**

A major focus in years two and three of the project has been to evolve specificity toward sequences identified by USAMRIID in two of the target toxins. Using an exploratory protease provided by Potomac (pS189), USAMRIID unambiguously identified the following cut sites:

| | |
|---|---|
| BoNT/B | FFMQ-S (exposed loop) |
| SEB | INSH-Q (exposed loop) |

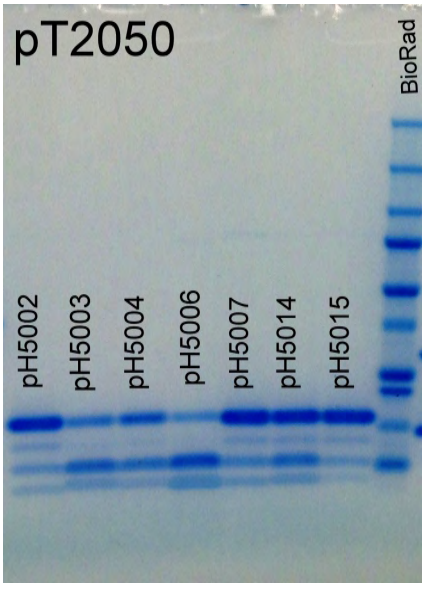
This effort primarily involved mutagenesis of subtilisin sub-sites S1-S7 and the use of phage display to select for mutants of desired specificity. Results are reported in detail in the Annual report for 2012. In year three we also used the information from phage-display selections to inform a computational design using Rosetta design software.

The two primary specificity pockets in subtilisin are the S1 and S4 site. The S1 pocket of subtilisin comprises amino acids, 127, 154, 156 and 166 and a water molecule that is hydrogen bonded to carbonyl oxygens of 126 and 152 and the main chain nitrogen of 169. The S4 site of subtilisin comprises amino acids at positions 104, 107, 126, 128, 130, 132 and 135. The natural preference of both S1 and S4 site are for hydrophobic amino acids (*11, 13, 17-20*).

Hydrophobic packing in both sub-sites is in some ways reminiscent of the protein folding problem. In the folding analogy, sub-site variation is viewed as mutation. Changes in P1 or P4 amino acid generally result in significant but not catastrophic losses in transition state stability. Among hydrophobic P1 amino acids, the $k_{cat}/K_M$ for P1 = Y is the best and P1 = A is 100-times less. $k_{cat}/K_M$ values for the remaining hydrophobic amino acids span the range in between. At the S4 sub-site, the preference for F relative to A is about 3- fold. (*19, 21*). A small P4 amino acid, such as alanine, points into the enzyme, but larger ones such as M, F, or Y, lie along a shallow indentation in the enzyme surface. The S4 pocket also has additional capacity, somewhat occluded behind the Tyr 104 residue. Y104 is able to adjust its position to accommodate larger or smaller amino acids.

In the protein folding analogy, a mutation in the hydrophobic core of a protein may decrease stability but is frequently not catastrophic because of adjustments in neighboring amino acids. To put the design problem into perspective, imagine designing a protein that is stably folded with one specific amino acid at a given position but unfolded with the other 19 amino acids at that position. This is obviously a much more challenging problem than just designing stabilizing or destabilizing mutations. This is basically what we would like to do in engineering protease specificity, however. Ideally one would like to engineer a sub-site so that only one amino acid supports catalysis. One way around this dilemma is to engineering disqualifying interactions at a sub-site – that is engineer interactions with non-cognate amino acids that are catastrophic. Steric clashes are one possible type of disqualifying interaction**.** In fact Van der Waals overlaps are the strongest non-covalent force associated with protein-protein interactions and create the possibility of decoding the binding of substrate amino acids that are too big to fit. Consequently we redesigned to S4 sub-site to try to uncode aromatic amino acids.

The original pT1001 mutant has an S4 site that is long but shallow. A shallow, solvent-accessible sub-site appears to promote P4 promiscuity. In a series of mutants, we close off part of the pocket to form a short, shallow pocket. This design was based on phage selections of mutants cleaving the sequence GRAL. Having identified a short, shallow pocket in selections, we then open up space in the interior of the S4 site. This space is excluded from solvent in a substrate complex, forming a deep, buried pocket for the P4 amino acid. To change the size and shape of the deep S4 pocket, we designed variations at amino acids 104, 107, 126, 128, 132 and 135. We have made these changes in combination with three different anion sites. This allows us to observe specificity in a series of mutants in which the acylation step becomes faster. In this series I30, P125 is the slowest, L30, P125 is moderate, and I30, S125 is fastest. In analyzing this series of variants, we note two general trends that are potentially useful. 1) Many different mutations at the sites 104, 107, 132 and 135 can be introduced without compromising high activity for certain P4 amino acids. These sites constitute a variable environment, with the effect of mutations largely isolated to effects on interactions with the P4 side chain. 2) Most mutations at some sites (e.g. 126, 128) decrease activity against all substrates. More than 100 random and site-directed variations were analyzed in the S4 engineering effort. The variant with the highest specificity for the target P4 specificity for SEB (P4=I) was pT2050. Kinetic results (Figure 3) with the closest P4 cognates amino acids shows the preference for P4 = I, followed by M and V. There is little activity vs. P4 = F (shown below), Y, or W (not shown).




F M A I L V T

**Figure 3**

Kinetic analysis was carried out with mutants list in *Appendix 6: P1 and P4 substrate series, pages 24-5.* S4 mutants were purified using *eglin variants listed in Appendix 7: Eglin variants, page 26.* A full list of selected S4 mutants subjected to kinetic analysis is shown in *Appendix 8: Co-evolution of S4 and Anion sites, pages 27-32.*

**Computational design of the S4 site for charged P4 amino acids.**

Another type of disqualifying interaction involves formation of an ion pair between a charged substrate amino acid and an oppositely charged amino acid in the binding pocket. The engineering challenge is that buried salt bridges are rare in nature and hard to engineer because the energy gained from the internal salt bridge must pay the cost of desolvation of the charged groups and also must compensate for lost interactions with counter-ions in solution. We had also previously failed in several attempts to evolve ionic interactions at the S4 sub-site by phage display. Using the knowledge of positions in the S4 site that allow for variation, we used Rosetta design software to generate numerous models for ion pair interactions. The Rosetta script use in the S4 design was:

```
104 S ALLAA
107 S ALLAA
108 S ALLAA
111 S ALLAA
122 S ALLAA
124 S ALLAA
132 S ALLAA
134 S ALLAA
135 S PIKAA RE          # R was used for P4 = E and E was used for P4 = K
139 S ALLAA
```

Six of these computationally designed mutants were expressed, purified and characterized: pT2114, pT2115, pT2121, pT2122, pT2123, pT2124 (*Appendix 8: Co-evolution of S4 and Anion sites, pages 31-32*). Two of these showed high specificity for a charged P4 amino acid: (P4 = E, pT2121 and P4 = K, pT2114).

**Engineering cooperative binding interactions at S1 and S4.**

Based on analysis of first generation phage selections and subsequent re-engineering by structure-based design, we believe that creating cooperativity between binding at S1 and S4 site has the potential to generate the highest specificity enzymes. The binding of a substrate to subtilisin appears to be a function of both the size and chemical complementarity of the side chain with a specific sub-site, as well as the global stability of the enzyme itself. The global enzyme stability comes into play because the beta strands comprising the peptide binding region can become distorted when destabilizing mutations are introduced even in distal regions of subtilsin. When a substrate binds, the beta strands reorganize into the canonical conformation. This reorganization is paid with substrate binding energy, weakening substrate binding. While this phenomenon complicates the interpretation of kinetic data, it can also potentially be exploited if substrate insertion and enzyme reorganization can be coupled in such a way as to cause cooperative binding interactions at sub-sites S1 and S4.

The S1 pocket, the S4 pocket and the anion site are all interconnected such that binding at one site can influence interactions at the other two. To promote this linkage we have mutated P168G. Proline at 168 is highly conserved in subtilisins and is in the rare cis conformation. By mutating this amino acid to glycine, we create space at the apex of the loop that forms the backs of the S1 and S4 sites and we

also destabilize the enzyme by replacing the rigid proline with the highly flexible glycine. This mutation was introduced into the backgrounds of S189 and S190 subtilisins. In these backgrounds, the mutation generally weakens substrate binding but has only a modest effect of specificity overall. A secondary effect is that the P168G mutation results in an amide proton deep in the S4 pocket, creating the potential for engineered polar interactions.

Other selected S1 and S4 mutants subjected to kinetic analysis are shown in *Appendix 8: Co-evolution of S4 and Anion sites, pages 27-32.*

**Use of engineered/evolved proteases in Protease Chain Reactions**

The central component of a synthetic ProCR is a self-amplifying complex. This is formed from a high-specificity, regulated subtilisin complexed with a high-affinity, but cleavable prodomain inhibitor. In describing the process, we will use the following terminology. "A" is a protease that cuts a cognate sequence "a". $I_a$ is a cleavable protease inhibitor that can be cut by protease A. $I_aA$ together form a self-amplifying complex. The protease is inactive when bound to the inhibitor but, once freed, is capable of cleaving the inhibitor and releasing additional free protease. This results in an exponentially expanding release of the active enzyme from the inactive complex until all subtilisin is liberated. The simplified mechanism of a protease chain reaction is $A + I_a A \rightarrow 2A$.

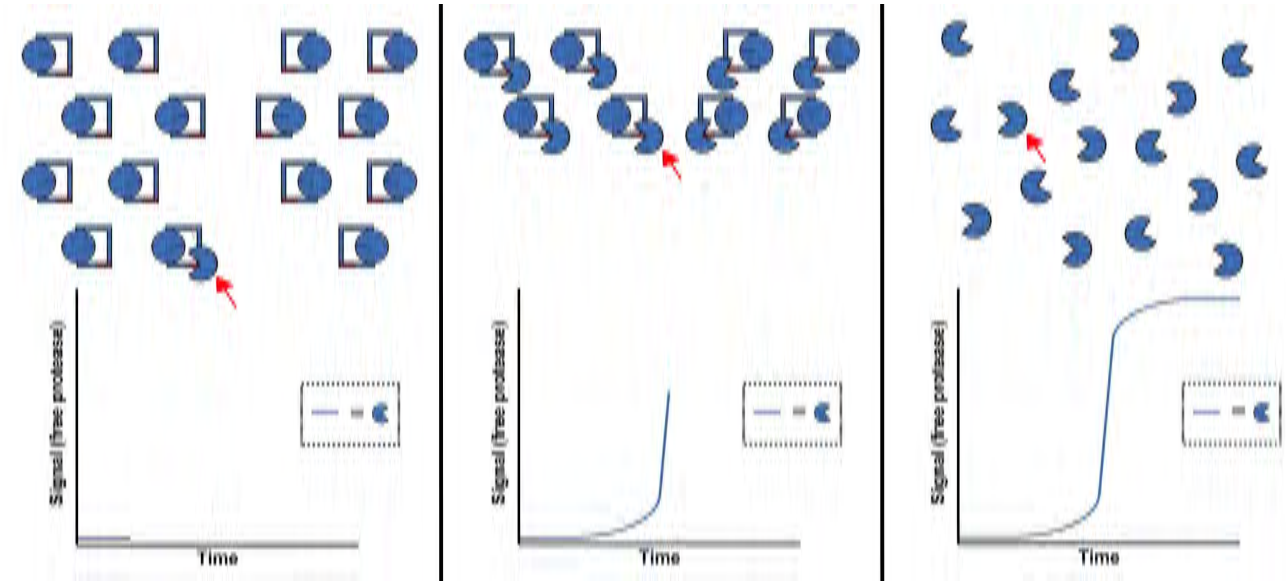



**Figure 4**

To create programmable cascades, three additional components were engineered. 1) Initiator proteases: an initiator protease ($A_0$) has the same specificity as its parent protease (A) but is not inhibited by the prodomain. 2) Incongruent complexes: incongruent complexes are formed from a protease "B" that cuts a cognate sequence "b" but not "a" such that $I_aB$ does not self-activate. Protease B can be released from the incongruent complex in the presence of a protease A, however. 3) Anti-inhibitors (denoted Q): an anti-inhibitor binds tightly to "I" but is itself inactive.

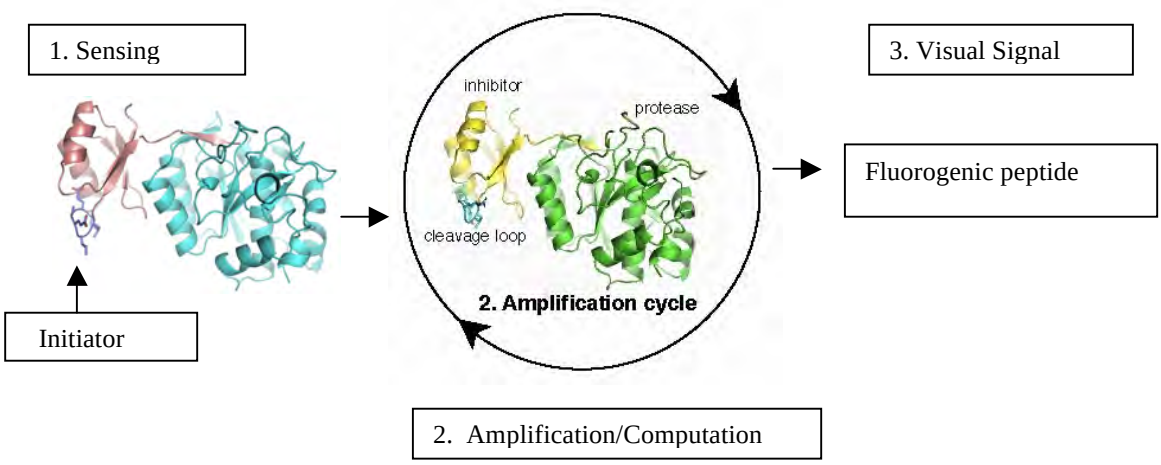



**Figure 5**

These components were assembly into activation cascades, and a mechanistic characterization of simple and compound cascades was carried out. Mechanistic understanding of chain reactions enables their use as programmable enzymatic sensors. Engineered protease chain reactions were able to measure concentrations of initiator protease 250 fM range in a 20 hour assay.

***Summary of progress on Statement of Work:***

Task 1: Chose cognate sequences from target toxins

1.1 The awardees shall review existing BoNT, SEB, ricin, and LF protein structures for amino acid sequences that present likely targets for RSUB. (Y1Q1)

*Completed*

Task 2: Evolve anion-regulated protease specificity

2.1 The awardees shall create a GA-COGNATE-GB phage capture protein for Task 1-identified target sites on each of the four toxins. (Y1Q3)

*Completed*

Also created GA-COGNATE-GB phage capture proteins with individual sub-site variations:

P2 = all twenty *complete*

P4 = all twenty *complete*

P1 + all twenty *complete*

2.2 The awardees shall create a phage library for each of the RSUB candidates in which the P1' and P2 anion-binding regions have been randomized. (Y1Q4)

*Completed*

Three anion libraries created and screened:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Library 1: | sites | 32 | 33 | 62 | 68 | 125 | |
| Library 2: | sites | 33 | 62 | 96 | 123 | 125 | 126 |

Library 3: sites 123 124 125 126 222 224 225

2.3 The awardees shall use phage display to identify library members which exhibit optimized anion-triggered GA-GB cleavage of selected toxin target sites. (Y2Q1)

*Completed*

Anion libraries screened:

| | | |
|---|---|---|
| Library 1 screened vs. | pH0101 | consensus sequence patterns obtained |
| Library 1 screened vs. | pH0106 | no consensus pattern obtained |
| Library 1 screened vs. | P2 = X | no consensus pattern obtained |
| Library 2 screened vs. | pH0101 | consensus sequence patterns obtained |

2.4 Starting from the four anion-optimized proteases evolved in 1.2.3, the awardees shall create phage libraries in which the P1 and P4 protease sites have been randomized. (Y2Q2)

*Completed*

Three P4 libraries created:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Library 1: | sites | 104 | 107 | 124 | 126 | 128 | |
| Library 2: | sites | 104 | 107 | 128 | 130 | 132 | 135 |
| Library 3 (optimized phagemid): | sites | 104 | 107 | 128 | 130 | 132 | 135 |

2.5 The awardees shall use phage display to select library members which exhibit the greatest specificity for each of the GA-GB capture proteins.(Y2Q3)

*Completed*

P4 libraries screened:

| | |
|---|---|
| Library 1 vs. P4 = A | consensus sequence patterns obtained |
| Library 1 vs. P4 = F | consensus sequence patterns obtained |
| Library 1 vs. P4 = I | mostly deletions mutants obtained: phagemid vector system optimized to control fusion protein expression |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Library 3 (optimized phagemid): | sites | 104 | 107 | 128 | 130 | 132 | 135 |
| Library 1 vs. P4 = G | consensus sequence patterns obtained | | | | | | |
| Library 1 vs. P4 = Q | consensus sequence patterns obtained | | | | | | |

2.6 Starting from the four anion-optimized proteases evolved in 1.2.5, the awardees shall create and screen phage libraries in which the P3 and P5 protease sites have been randomized. (Y3Q1)

*Completed*

Random and designed mutations analyzed for P1, P3, P5 and P7 protease sites: Positions 108, 109, 112, 128, 129, 133, 134, 137, 144, 152, and 166. The specific mutants analyzed are listed in Appendices 4 and 8.

Task 3: Characterize catalytic properties of engineered proteases.

*Completed*

3.1 The awardees shall use subtilisin-Alexafluor 350 conjugates to measure protease kinetics with substrates containing each of the cognate toxin sequences. (Y3Q2)

*Completed*

Kinetics analysis performed with key mutants.

Promising mutants given to USAMRIID for testing with toxins.

Protease chain reaction assay developed to assay activity and specificity of sequences in a structured environment.

## KEY RESEARCH ACCOMPLISHMENTS IN YEAR 3

1. Determined the structure of an evolved variant pT2077 in complex with the substrate sequence used to select it.
2. Design/evolution of a highly active enzyme that can cut P4 = I (pT2050).
3. Computational design of specificity for an ionic P4 amino acid (P4 = E, pT2121 and P4 = K, pT2114);
4. Engineered protease chain reactions that can reliably measure concentrations of 250 fM range in a 20 hour assay.

## REPORTABLE OUTCOMES

Presented posters at 2011 and 2012 Spring Research Festivals
Presented posters 2011ASM general meeting
Present a talk at the 2011 DTRA Biodefense Conference
Bryan, P. N. (2012) *Engineering Protease Specificity*, in The Protein Engineering Handbook Vol. III, Lutz and Bornscheuer, eds., Wiley Press, Weinheim. (pp 243-278).
Coordinates of 1.3Å x-ray structure for pT2077 to be deposited in Protein Data Bank.

## CONCLUSIONS

General conclusions concerning protease engineering are described in Bryan, P. N. (2012) *Engineering Protease Specificity*. Broader implications of results on enzyme engineering are discussed below.

Engineering a custom catalyst for an arbitrary chemical reaction remains a difficult challenge. If enzyme engineering is viewed as creating components that can be assemble into more complex machines, however, the task becomes tractable. A by-product of this work is the demonstration that complex enzymatic machines can be constructed based on simpler, well-understood component parts. Serine proteases and their inhibition have been studied for decades and offer unique opportunities for re-purposing into enzymatic machines (*1, 2, 4, 22-24*). For example, we previously developed an anion-triggered subtilisin that was combined with a prodomain tag to create simple methodology to affinity purify recombinant proteins and remove the affinity tag in one step (Profinity eXact System, Bio-

Rad) (*16*). The "switchable" subtilisin used for protein purification is a general component that can be applied to numerous enzymatic problems. For example, here we used the purification function of a switchable protease to create a selection system to further evolve enzymatic function. The purification function was used to parse random sequence space and "purify" protease variants that cleave a specific sequence in response to specific anions such as azide or nitrite.

Another powerful property of proteases is their ability to self-activate, self-amplify, and propagate signals when bound to certain protein inhibitors. In fact, natural protease cascades regulate cellular processes from embryogenesis to cell death by linking diverse enzymatic functions together with multiple logic gates, e.g. (*25*). The engineered elements described here (sequence-specific proteases, cleavable inhibitors, and small-molecule activators) were used to build and characterizing synthetic cascades. These synthetic cascades were developed initially to characterize proteases evolved in this project, but were subsequently reprogrammed into molecular sensors.

**Molecular sensors**

A sensor consists of a detector that responds to an analyte, the ProCR that amplifies and quantifies the response from the detector, and a transducer that produces a signal. Through this combination, extremely sophisticated enzymatic sensors can be built, powered by only the chemical energy of the constituent enzymes.

***Detector element*** ProCR can be used to detect any analyte that perturbs the initial protease-inhibitor equilibrium. There are many variations of this basic idea. Three are briefly described below.

*1. Anion detection* Because certain anions increase the rate of loop cleavage in the inhibitor, activating anions can be detected. Detection of azide is illustrated here but other anions of interest can also be measured by using other protease variants in self-amplifying cascades. These include hydroxide (pH), fluoride, and nitrite. Nitrite is an indicator of many disease states, as it is a stable oxidation product of the short-lived, signaling molecule nitric oxide. Azide is present in some high explosives and fluoride is a breakdown product of nerve agents, such as Soman and Sarin.

*2. Linkage to binding molecules* Conjugating an antibody (or any other binding module) with an initiating protease or an anti-inhibitor allows a cascade to be incorporated into virtually any immunoassay to improve its sensitivity and ability to precisely measure the concentration of the target molecule. For example, in the ProCR version of an enzyme-linked immunosorbent assay (ELISA), the target protein is immobilized on a solid support, an antibody conjugated to an initiating protease forms a complex with the target protein, and then self-amplifying complex and substrate are added to amplify the signal from the conjugated protease and convert its concentration into a time signature.

*3. Protease detection* Incongruent complexes coupled to self-amplifying complexes in compound cascades can be used to detect and quantify the presence of any protease with a well-defined substrate specificity. A sensor complex contains the cognate sequence of a natural protease in the loop (**Fig. 5**). The sensor complex is not self-amplifying but the proteolytic action of the target protease releases free subtilisin protease by cutting the exposed loop on the detection inhibitor. The free subtilisin, in turn, initiates a self-amplifying chain reaction. This has important implications for clinical diagnostics because proteases are already widely-used biomarkers. Examples include granzymes, matrix metalloproteases (MMPs) and kallikreins (KLKs), which includes the prostate-specific antigen, KLK3. A major challenge is that assays for single proteases often lack the sensitivity and specificity to be clinically useful. As the protease sensor technology develops, we should be able to detect multiple proteases and perform multiparametric analysis of protease marker panels. The more complex the enzymatic machinery, the more powerful the diagnostic capability will be. Protease sensors can also be used for the detection of pathogens by sensing the specific proteases they produce (e.g. *Bacillus anthracis* lethal factor and *Botulinum* neurotoxin A). Host proteases can also be monitored as indicators of infection. In general, we would like to develop the core technology to detect anything that produces a specific protease or any physiological event that causes specific proteases to be produced in response.

***Processing/computational element*** A ProCR is a powerful analog computer with two characteristics that greatly facilitate the detection of analytes. 1) It can convert the concentration of an initiating analyte into a time signature. 2) It can create enormous signal amplification, analogous to the amplification of DNA by PCR. Thus detection is enabled because the final observable signal can be very large and the time lag until onset of the signal is precisely correlated with the concentration of initiating analyte. How the reaction responds to the detector element is determined by numerous adjustable parameters including the concentrations of self-amplifying complex, free inhibitor, triggering anions, and buffer salts.

Combining different types of complexes (incongruent, anti-inhibitor, ramping, and self-amplifying) into compound cascades creates tremendous versatility in maximizing the response to a target analyte and minimizing the background response. Mechanistic understanding is critical in designing useful compound cascades. The response curves of compound cascades are very reliable, but all relevant equilibria must be well-defined. In the absence of mechanistic understanding, the kinetic response is usually not intuitive. The individual equilibria are like lines of code within a larger chemical program. The relationship between target molecule concentration and the lag time is chemically programmed into inhibitor-protease pairs. Different binding constants and kinetic parameters in the binding, cleavage, and release steps result in different responses to target molecule. Relevant equilibria include not only the inhibition by the intact inhibitor and kinetic parameters for cleavage loop, but also inhibition by all inhibitor fragments and substrate products. Non-native interactions between the protease and the inhibitor must also be ruled out for any given set of components. The high sensitivity of multi-component enzymatic cascades to small variations is a challenge to their characterization but is the key to their utility.

***Transduction element*** The transducer in ProCR can be anything that is changed by the protease released in a self-amplifying reaction. Signaling is one example, but the protease released can also mediate other outputs. By being able to activate or inactivate other proteins, a protease can act as a transistor in an enzymatic circuit. Simple components, once fully characterized, can be combined to form switches, signal amplifiers, and transducers. Note that proteases are particularly useful enzymes to incorporate into enzymatic machines because, in addition to generating optical signals, they can also modify other proteins in reaction cascades.

**So what?** If one considers the construction of sophisticated electronic devices from standard components, one can appreciate the enormous potential of creating enzymatic machines from standard components that link diverse enzymatic function. As the technology develops, engineered proteases can be used for increasingly complex functions, such as measuring and controlling cellular processes. This has implications for biodefense because ProCR may eventually be used to detect the molecular signature of a pathogen, as well as produce a specific therapeutic response.

## REFERENCES


(1) Perona, J. J., and Craik, C. S. (1995) Structural basis of substrate specificity in the serine proteases. *Protein Sci 4*, 337-360.

(2) Hedstrom, L. (2002) Serine protease mechanism and specificity. *Chem Rev 102*, 4501-4524.

(3) Wells, J. A. (1990) Additivity of mutational effects in proteins. *Biochemistry 29*, 8509-8517.

(4) Radisky, E. S., Kwan, G., Karen Lu, C. J., and Koshland, D. E., Jr. (2004) Binding, Proteolytic, and Crystallographic Analyses of Mutations at the Protease-Inhibitor Interface of the Subtilisin BPN'/Chymotrypsin Inhibitor 2 Complex(,). *Biochemistry 43*, 13648-13656.

(5) Otzen, D. E., and Fersht, A. R. (1999) Analysis of protein-protein interactions by mutagenesis: direct versus indirect effects. *Protein Eng 12*, 41-45.

(6) Lu, W., Apostol, I., Qasim, M. A., Warne, N., Wynn, R., Zhang, W. L., Anderson, S., Chiang, Y. W., Ogin, E., Rothberg, I., Ryan, K., and Laskowski, M., Jr. (1997) Binding of amino acid side-chains to S1 cavities of serine proteinases. *J Mol Biol 266*, 441-461.

(7) Eder, J., Rheinnecker, M., and Fersht, A. R. (1993) Hydrolysis of small peptide substrates parallels binding of chymotrypsin inhibitor 2 for mutants of subtilisin BPN'. *FEBS Lett 335*, 349-352.

(8) Berger, A., and Schechter, I. (1970) Mapping the active site of papain with the aid of peptide substrates and inhibitors. *Philos Trans R Soc Lond B Biol Sci 257*, 249-264.

(9) McPhalen, C. A., and James, M. N. G. (1988) Structural comparison of two serine proteinase-protein inhibitor complexes: Eglin-C-Subtilisin Carlsberg and CI-2-Subtilisin novo. *Biochemistry 27*, 6582-6598.

(10) McPhalen, C. A., Schnebli, H. P., and James, M. N. (1985) Crystal and molecular structure of the inhibitor eglin from leeches in complex with subtilisin Carlsberg. *FEBS Lett 188*, 55-58.

(11) Estell, D. A., Graycar, T. P., Miller, J. V., Powers, D. B., Burnier, J. P., Ng, P. G., and Wells, J. A. (1986) Probing steric and hydrophobic effects on enzyme-substrate interactions by protein engineering. *Science 233*, 659-663.

(12) Gron, H., and Breddam, K. (1992) Interdependency of the binding subsites in subtilisin. *Biochemistry 31*, 8967-8971.

(13) Gron, H., Meldal, M., and Breddam, K. (1992) Extensive comparison of the substrate preferences of two subtilisins as determined with peptide substrates which are based on the principle of intramolecular quenching. *Biochemistry 31*, 6011-6018.

(14) Gallagher, D. T., Ruan, B., London, M., Bryan, M. A., and Bryan, P. N. (2009) Structure of a switchable subtilisin complexed with substrate and with the activator azide. *Biochemistry 48* 10389-10394.

(15) Johnson, K. A. (1992) Transient-state kinetic analysis of enzyme reaction pathways. *The Enzymes 20*, 1-61.

(16) Ruan, B., Fisher, K. E., Alexander, P. A., Doroshko, V., and Bryan, P. N. (2004) Engineering subtilisin into a fluoride-triggered processing protease useful for one-step protein purification. *Biochemistry 43*, 14539-14546.

(17) Sorensen, S. B., Bech, L. M., Meldal, M., and Breddam, K. (1993) Mutational replacements of the amino acid residues forming the hydrophobic S4 binding pocket of subtilisin 309 from Bacillus lentus. *Biochemistry 32*, 8994-8999.

(18) Bech, L. M., Sorensen, S. B., and Breddam, K. (1993) Significance of hydrophobic S4-P4 interactions in subtilisin 309 from Bacillus lentus. *Biochemistry 32*, 2845-2852.

(19) Rheinnecker, M., Baker, G., Eder, J., and Fersht, A. R. (1993) Engineering a novel specificity in subtilisin BPN'. *Biochemistry 32*, 1199-1203.

(20) Rheinnecker, M., Eder, J., Pandey, P. S., and Fersht, A. R. (1994) Variants of subtilisin BPN' with altered specificity profiles. *Biochemistry 33*, 221-225.

(21) Gron, H., Bech, L. M., Sorensen, S. B., Meldal, M., and Breddam, K. (1996) Studies of binding sites in the subtilisin from Bacillus lentus by means of site directed mutagenesis and kinetic investigations. *Adv Exp Med Biol 379*, 105-112.
(22) Carter, P., and Wells, J. A. (1988) Dissecting the catalytic triad of a serine protease. *Nature 332*, 564-568.
(23) Wells, J. A., and Estell, D. A. (1988) Subtilisin--an enzyme designed to be engineered. *Trends Biochem. Sci. 13*, 291-297.
(24) Sauer, R. T., and Baker, T. A. (2011) AAA+ proteases: ATP-fueled machines of protein destruction. *Annu Rev Biochem 80*, 587-612.
(25) Danial, N. N., and Korsmeyer, S. J. (2004) Cell death: critical control points. *Cell 116*, 205-219.

## Appendix 1: pHen vectors

| Name | Parent | Vector description | Cloned subtilisin description |
|---|---|---|---|
| phenT924 | phen924 | phenT=phen vector with tac instead of lac promoter | S189 |
| phenT926 | phenT924 | phenT=phen, tac | anion mut.1+T166 (as in pT1001) |
| phenT928 | phenT926 | phenT=phen, tac | anion mut.1+T166, Q10 replaced with amber codon, linker's amber codon replaced with Q |
| phenT929 | phenT926 | phenT=phen, tac | anion mut.1+T166, two amber codons: linker and subtilisin(Q10 position) |
| phen927 | phen926 | phen | anion mut.1+T166, no amber codons |
| phen928 | phen926 | phen | anion mut.1+T166, Q10 replaced with amber codon, linker's amber codon replaced with Q |
| phen929 | phen926 | phen | anion mut.1+T166, two amber codons: linker and subtilisin(Q10 position) |
| phen926 | phen924 | phen | anion mut.1+T166 (as in pT1001) |
| phen/srp928 | phenT928 | phenT=phen, srp promoter, Natasha's lib. Book#2, p.55<br>srp promoter: spacer between -35 and -10 of tac promoter<br>replaced with lac operator, original seq of lacO is removed. | anion mut.1+T166, Q10 replaced with amber codon, linker's amber codon replaced with Q |

## Appendix 2: Primary Anion screening

| Strain name | Parent strain | 30 | 32 | 33 | 62 | 68 | 104 | 105 | 107 | 125 | 134 | 166 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **pN0001-pN0330 DO NOT have XhoI site.** | | | | | | | | | | | | |
| pN0001 | **anion library screen 1mM NaNO2 5mins.** | I | G | T | S | I | | | | P | | |
| pN0003 | | I | G | T | N | L | | | | P | | |
| pN0005 | | I | S | S | N | T | | | | S | | |
| pN0007 | | I | G | T | N | I | | | | P | | |
| pN0008 | | I | G | T | N | I | | | **V** | P | | |
| pN0016 | | I | G | T | N | S | | | | P | | |
| pN0020 | | L | G | T | N | M | | | | P | | |
| pN0021 | | I | G | T | R | M | | | | S | | |
| pN0027 | | L | G | T | N | A | | | | P | | |
| pN0308 | **anion library screen 10mM NaNO2 5mins.** | I | G | T | A | I | | | | P | | |
| pN0310 | | V | S | T | A | C | | | | S | | |
| pN0313 | | L | G | S | G | L | | | | A | | |
| pN0320 | | I | S | T | N | I | | | | S | | |
| pN0321 | | I | G | T | L | A | | | | S | | |
| pN0322 | | L | G | T | N | S | | | | P | | |
| pN0323 | | V | S | T | N | T | | | | S | | |
| pN0327 | | V | G | T | N | A | | | | P | | |
| pN0330 | | L | S | T | N | A | | | | P | | |

## Appendix 3: Secondary anion screening with variable P2

| Strain name | Parent strain | 30 | 32 | 33 | 62 | 68 | 125 |
|---|---|---|---|---|---|---|---|
| pT0401 | pTac vector+PCR product from anion library screen with 1mM Azide 5mins on FRTL. | L | S | T | N | G | P |
| pT0402 | | V | S | T | N | E | P |
| pT0403 | | L | T | G | K | V | T |
| pT0404 | | I | G | P | Q | G | P |
| pT0405 | | L | S | T | N | G | S |
| pT0406 | | L | S | T | N | S | P |
| pT0407 | | V | T | A | T | G | D |
| pT0408 | | I | G | T | N | G | S |
| pT0409 | | L | G | S | S | V | Q |
| pT0410 | | I | S | N | T | G | N |
| pT0411 | pTac vector+PCR product from anion library screen with 1mM Azide 5mins on FRLL. | I | A | N | N | A | P |
| pT0412 | | I | A | G | I | A | V |
| pT0413 | | T | G | S | A | N | P |
| pT0414 | | I | S | S | T | S | D |
| pT0415 | | V | S | T | N | L | D |
| pT0416 | | L | G | G | L | Q | Q |
| pT0417 | | V | G | S | L | A | Y |
| pT0418 | | V | S | T | N | E | T |
| pT0419 | pTac vector+PCR product from anion library screen with 1mM Azide 5mins on FRGL | M | G | Y | S | A | P |
| pT0420 | | V | S | T | V | V | N |
| pT0421 | | L | S | T | N | Q | P |
| pT0422 | pTac vector+PCR product from anion library screen with 1mM Azide 5mins on FREL | L | T | N | T | A | P |
| pT0423 | | I | G | G | L | T | S |
| pT0424 | | I | S | S | T | A | P |
| pT0425 | | L | G | T | N | Q | T |
| pT0426 | | L | D | G | G | S | G |
| pT0427 | | M | G | T | N | E | N |
| pT0428 | | I | S | S | T | S | P |
| pT0429 | | I | G | G | D | D | S |
| pT0430 | | L | S | S | L | A | T |
| pT0431 | | I | A | T | L | A | P |
| pT0432 | | V | S | T | A | Q | P |
| pT0433 | | L | G | E | T | L | N |
| pT0434 | pTac | I | S | T | L | M | T |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| pT0435 | **vector+PCR product from anion library screen with 1mM Azide 5mins on FRRL.** | I | G | T | N | Q | S |
| pT0436 | | I | S | T | N | S | P |
| pT0437 | | M | G | P | T | D | S |
| pT0438 | | I | S | S | T | M | G |
| pT0439 | | L | S | T | N | M | P |
| pT0440 | | I | G | T | N | T | D |
| pT0441 | | L | S | S | Y | I | P |
| pT0442 | | V | A | D | S | A | P |
| pT0443 | | S | S | S | L | L | S |
| pT0444 | | M | G | G | S | A | D |
| pT0445 | | V | G | T | S | N | N |
| pT0446 | | I | S | A | T | M | N |

## Appendix 4: Tertiary anion library with variable P1

| Strain name | Nick name | 30 | 32 | 33 | 62 | 68 | 123 | 124 | 125 | 126 | 152 | 166 | 222 | 224 | 225 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | V | A | S | N | V | N | M | S | L | A | S | M | S | P |
| **pT1000** | **pTan1D166** | I | G | T | S | I | | | P | | A | D | | | |
| pT1001 | pTan1T166 | I | G | T | S | I | | | P | | A | T | | | |
| pT1002 | pT2004cystail | | | | | | | | | | | | | | |
| pT1003 | | | | | | | | | | | | | | | |
| pT1004 | pT1001 123/H | I | G | T | S | I | H | | P | | A | T | | | |
| pT1005 | pT1001 126/Y | I | G | T | S | I | | | P | Y | A | T | | | |
| pT1006 | pT1001 123A 126Y | I | G | T | S | I | A | | P | Y | A | T | | | |
| pT1007 | pT1001 123C 126T | I | G | T | S | I | C | | P | T | A | T | | | |
| pT1008 | pT1001 123L 126Y | I | G | T | S | I | L | | P | Y | A | T | | | |
| pT1009 | pT1001 12H 126F | I | G | T | S | I | H | | P | F | A | T | | | |
| pT1010 | mut6B | | | | | | G | L | P | I | | | I | L | G |
| pT1011 | ptacs170 | | | | | | | | | | | | | | |
| pT1012 | ptacpro-s189 | | | | | | | | | | | | | | |
| pT1013 | 30sec lib.mutant | | | | | | Y | V | S | A | | | L | V | A |
| pT1014 | 30sec lib.mutant | | | | | | L | T | M | L | | | T | S | A |
| pT1015 | 30sec lib.mutant | | | | | | N | M | P | L | | | T | Q(TAG) | A |
| pT1016 | 30sec lib.mutant | | | | | | N | M | P | L | | | T | Q(TAG) | G |
| pT1017 | 30sec lib.mutant | | | | | | N | M | P | L | | | R | S | S |
| pT1018 | 30sec lib.mutant | | | | | | A | V | P | L | | | R | V | L |
| pT1019 | 30sec lib.mutant | | | | | | L | S | T | Y | | | T | I | S |
| pT1020 | 30sec lib.mutant | | | | | | Y | V | S | A | | | L | V | A |
| pT1021 | | | | | | | N | M | P | L | | | T | Q(CAG) | A |
| pT1022 | | | | | | | N | M | P | L | | | T | Q(CAG) | G |
| pT1023 | T8A | | | | | | A | T | H | I | | | P | Y | A |
| pT1024 | T11A | | | | | | N | M | P | L | | | R | S | S |
| pT1026 | pT1010QC2 | | | | | | G | L | P | I | | | M | L | G |
| pT1027 | pT1018QC3, 4 | | | | | | A | V | P | L | | | M | V | L |
| pT1028 | pT1022QC1, 2 | | | | | | N | M | P | L | | | T | L | A |

## Appendix 5: Substrate proteins with variable P2

| Strain name | Nick name | | Parent strain | P5 | P4 | P3 | P2 | P1 | P1' | P2' |
|---|---|---|---|---|---|---|---|---|---|---|
| pH0101 | 6002 | | Ga(P220)proGb1 6His(FKAM_SG) | V | F | K | A | M | S | G |
| pH0102 | 6002 mut2 | | | V | Y | R | A | L | S | G |
| pH0103 | 6002 mut3 | | | V | F | G | V | L | D | E |
| pH0104 | 6002 mut4 | | | V | Y | K | I | F | T | G |
| pH0105 | 6002 mut5 | | | V | F | S | A | Y | K | K |
| pH0106 | 6002 mut6 | | | V | F | T | G | L | M | E |
| pH0107 | 6002 mut7 | | | V | Y | Y | E | M | S | G |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | **VFRXL-SG** | | | | | | | |
| pH0201 | 6002 mut1 - 1 | P2 pro substrate FRXL | **all are from 6002 mut1. ( VFRAL-SG)** | V | F | R | T | L | S | G |
| pH0202 | 6003 mut1 - 2 | | | V | F | R | L | L | S | G |
| pH0203 | 6002 mut1 - 3 | | | V | F | R | A | L | S | G |
| pH0204 | 6002 mut1 - 4 | | | V | F | R | S | L | S | G |
| pH0205 | 6002 mut1 - 5 | | | V | F | R | M | L | S | G |
| pH0206 | 6002 mut1 - 8 | | | V | F | R | Q | L | S | G |
| pH0207 | 6002 mut1 - 10 | | | V | F | R | V | L | S | G |
| pH0208 | 6002 mut1 - 13 | | | V | F | R | Y | L | S | G |
| pH0209 | 6002 mut1 - 16 | | | V | F | R | E | L | S | G |
| pH0210 | 6002 mut1 - 17 | | | V | F | R | C | L | S | G |
| pH0211 | 6002 mut1 - 20 | | | V | F | R | K | L | S | G |
| pH0212 | 6002 mut1 - 23 | | | V | F | R | G | L | S | G |
| pH0213 | 6002 mut1 - 26 | | | V | F | R | P | L | S | G |
| pH0214 | 6002 mut1 - 56 | | | V | F | R | W | L | S | G |
| pH0215 | 6002 mut1 - 58 | | | V | F | R | I | L | S | G |
| pH0216 | 6002 mut1 - 59 | | | V | F | R | R | L | S | G |
| pH0217 | 6002 mut1 - 43 | | | V | F | R | H | L | S | G |
| pH0218 | 6002 mut1 - 60 | | | V | F | R | D | L | S | G |
| pH0219 | 6002 mut1 - 62 | | | V | F | R | F | L | S | G |
| pH0220 | 6002 mut1 - 15 | | | V | F | R | N | L | S | G |

## Appendix 6: P1 and P4 substrate series

| Strain name | | P5 | P4 | P3 | P2 | P1 | P1' | P2' | P3' | P4' | P5' |
|---|---|---|---|---|---|---|---|---|---|---|---|
| pH5001 | P4 no pro substrate XRAL | L | F | R | A | L | S | A | T | G | T |
| pH5002 | | L | F | R | A | L | M | A | K | G | T |
| pH5003 | | L | M | R | A | L | M | A | K | G | T |
| pH5004 | | L | A | R | A | L | M | A | K | G | T |
| pH5005 | | L | G | R | A | L | M | A | K | G | T |
| pH5006 | | L | I | R | A | L | M | A | K | G | T |
| pH5007 | | L | L | R | A | L | M | A | K | G | T |
| pH5008 | | L | C | R | A | L | M | A | K | G | T |
| pH5009 | | L | P | R | A | L | M | A | K | G | T |
| pH5010 | | L | H | R | A | L | M | A | K | G | T |
| pH5011 | | L | K | R | A | L | M | A | K | G | T |
| pH5012 | | L | S | R | A | L | M | A | K | G | T |
| pH5013 | | L | W | R | A | L | M | A | K | G | T |
| pH5014 | | L | V | R | A | L | M | A | K | G | T |
| pH5015 | | L | T | R | A | L | M | A | K | G | T |
| pH5016 | | L | R | R | A | L | M | A | K | G | T |
| pH5017 | | L | E | R | A | L | M | A | K | G | T |
| pH5018 | | L | N | R | A | L | M | A | K | G | T |
| pH5019 | | L | D | R | A | L | M | A | K | G | T |
| pH5020 | | L | Y | R | A | L | M | A | K | G | T |
| pH5021 | | L | Q | R | A | L | M | A | K | G | T |
| pH5022 | | L | F | R | A | L | M | A | K | S | S |
| | | | | | | | | | | | |
| pH5050 | no pro substrate FRAX | L | F | R | A | H | M | A | K | G | T |
| pH5051 | | L | F | R | A | S | M | A | K | G | T |
| pH5052 | | L | F | R | A | L | M | A | K | G | T |
| pH5053 | | L | F | R | A | G | M | A | K | G | T |
| pH5054 | | L | F | R | A | R | M | A | K | G | T |
| pH5055 | | L | F | R | A | N | M | A | K | G | T |
| pH5056 | | L | F | R | A | M | M | A | K | G | T |
| pH5057 | | L | F | R | A | Q | M | A | K | G | T |
| pH5058 | | L | F | R | A | E | M | A | K | G | T |
| pH5059 | | L | F | R | A | A | M | A | K | G | T |
| pH5060 | | L | F | R | A | T | M | A | K | G | T |
| pH5061 | | L | F | R | A | C | M | A | K | G | T |
| pH5062 | | L | F | R | A | V | M | A | K | G | T |
| pH5063 | | L | F | R | A | P | M | A | K | G | T |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| pH5064 | | L | F | R | A | F | M | A | K | G | T |
| pH5065 | | L | F | R | A | D | M | A | K | G | T |
| pH5066 | | L | F | R | A | I | M | A | K | G | T |
| pH5067 | | L | F | R | A | K | M | A | K | G | T |
| pH5068 | | L | F | R | A | W | M | A | K | G | T |
| pH5069 | | L | F | R | A | Y | M | A | K | G | T |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| pH5101 | | L | F | R | A | L | S | A | T | G | T |
| pH5102 | | L | F | R | A | L | S | A | T | G | T |
| pH5103 | | L | X | R | A | X | S | A | T | G | T |
| pH5104 | | L | P | R | A | T | S | A | T | G | T |
| pH5105 | | L | E | R | A | L | S | A | T | G | T |
| pH5106 | | L | F | R | *(TAA) | L | S | A | T | G | T |
| pH5107 | | L | Q | R | A | L | S | A | T | G | T |
| pH5108 | | L | P | R | A | L | S | A | T | G | T |
| pH5109 | | L | Q | R | A | L | | | | | |
| pH5110 | | L | L | R | A | L | | | | | |
| pH5111 | | L | G | R | A | L | | | | | |
| pH5112 | | L | A | R | A | L | | | | | |
| pH5113 | | L | D | R | A | L | | | | | |
| pH5114 | | L | T | R | A | L | | | | | |
| pH5115 | | L | N | R | A | L | | | | | |

## Appendix 7: Eglin vectors

| Plasmid | Nickname | Parent Strain | Amino Acid changes | | | | |
|---|---|---|---|---|---|---|---|
| | | | # | 20 | 23 | 27 | 42 |
| | | | pG48 | Q | E | L | P |
| pG48 | | pG5 | | Q | E | L | P |
| pK1 | P42YA | pG48 | | | | | Y |
| pK2 | A2 | pG48 | | | | C | |
| pK3 | B4 | pG48 | | C | C | C | |
| pK4 | C3 | pG48 | | K aag | K aag | K aag | |
| pK5 | 260E | pG48 | | K aaa | K aaa | K aaa | |
| | | | | | | | |
| | | | | | | | |
| pK6 | 71721, 71722, 71724 | pK1 | | | | | |
| pK7 | 34A | pK4 | | K aag | K aag | K aag | F |
| pK8 | 56A | pK4 | | K aag | K aag | K aag | D |
| pK9 | 78A | pK4 | | K aag | K aag | K aag | E |
| pK10 | 90A | pK4 | | K aag | K aag | K aag | K |

**Appendix 8: Co-evolution of P4 and Anion sites**

| Strain name | Parent strain | 30 | 32 | 33 | 62 | 68 | 104 | 105 | 107 | 108 | 125 | 128 | 129 | 130 | 132 | 134 | 135 | 166 | 168 | 221 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | V | A | S | N | V | A | S | I | I | S | S | P | S | S | A | L | S | P | S |
| pT2001 | ptac vector | | S | | | | | | | | | | | | | | | | | |
| pT2002 | pT2001 | | S | | | | | | | | | | | | | | | | | G |
| pT2003 | pT2002 | | D | | | | | | | | | | | | | | | | | G |
| pT2004 | pT2001 | | A | | | | | | | | | | | | | | | | | |
| pT2005 | ptac vector (924 library) polarXRAL | | | | | | C | | D | | | R | | K | P | | Y | | | |
| pT2006 | | | | | | | Q | | N | | | N | | W | S | | T | | | |
| pT2007 | | | | | | | A | | I | | | A | | S | S | | L | | | |
| pT2008 | pT0001 | I | G | T | S | I | Y | | | | P | | | | | | | | | |
| pT2009 | pT0001 | I | G | T | S | I | Y | | | | P | G | | | | | | | | |
| pT2010 | pT2004 | | | | | | Y | | | | | | | | | | | | | |
| pT2011 | pT2004 | | | | | | Y | | | | | G | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
| | pHENT928 | I | G | T | S | I | A | | I | 108 | P | S | | S | S | | L | T | | |
| pT2012 | ptac vector (pHENT928 library) GRAL substrate | I | G | T | S | I | A | | I | | P | L | | Q | V | | L | T | | |
| pT2013 | | I | G | T | S | I | A | | I | | P | F | | E | S | | V | T | | |
| pT2014 | | I | G | T | S | I | A | | I | | P | I | | S | S | | L | T | | |
| pT2015 | | I | G | T | S | I | A | | I | | P | V | | S | S | | L | T | | |
| pT2016 | | I | G | T | S | I | A | | I | | P | V | | G | S | | L | T | | |
| pT2017 | | I | G | T | S | I | A | | I | | P | F | | G | S | | L | T | | |
| pT2018 | | I | G | T | S | I | A | | I | | P | L | | G | H | | L | T | | |
| pT2019 | | I | G | T | S | I | A | | I | | P | I | | T | S | | L | T | | |
| pT2020 | | I | G | T | S | I | A | | I | | P | L | | G | Q | | L | T | | |
| pT2021 | | I | G | T | S | I | A | | I | | P | L | | D | S | | L | T | | |
| | | | | | | | | | | | | | | | | | | | | |
| pT2022 | ptac vector (pHENT928 library) QRAL substrate | I | G | T | S | I | T | | V | | P | S | | S | S | | L | T | | |
| pT2023 | | I | G | T | S | I | A | | I | | P | P | | N | D | | A | T | | |
| | | I | G | T | S | I | T | | L | | P | S | | S | S | | L | T | | |
| | | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| pT2024 | **pT2014** | **L** | G | T | S | I | A | | I | | **S** | I | | S | S | | L | T | | |
| pT2025 | **pT2014** | **L** | G | T | S | I | A | | I | | P | I | | S | S | | L | T | | |
| pT2026 | **pT2014** | I | G | T | S | I | A | | I | | P | I | | S | S | | L | T | | |
| | | | | | | | | | | | | | | | | | | | | |
| pT2027 | **pT2004+E112A** | | | | | | | | | | | | | | | | | | | |
| pT2028 | **pT2004+A134K** | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |
| | | 30 | 32 | 33 | 62 | 68 | 104 | 105 | 107 | 108 | 125 | 128 | 129 | 130 | 132 | 134 | 135 | 166 | 168 | 221 |
| pT2029 | **pT2026** | I | G | T | S | I | A | | I | | **S** | I | | S | S | A | L | T | | |
| pT2030 | **pT2029** | I | G | T | S | I | A | | I | | S | I | | S | S | K | L | T | | |
| pT2031 | **pT2029** | I | G | T | S | I | A | | I | | S | S | | S | S | A | L | T | | |
| pT2032 | **pT2030** | I | G | T | S | I | A | | I | | S | S | | S | S | K | L | T | | |
| pT2033 | **pT2031** | I | G | T | S | I | A | C | I | | S | S | | S | S | A | L | T | | |
| pT2034 | **pT2031** | I | G | T | S | I | A | | I | | S | S | | S | S | C | L | T | | |
| pT2035 | **pT2026** | I | G | T | S | I | T | | L | | P | I | | S | S | | L | T | | |
| pT2036 | **pT2035** | second site pro cleavage AKAL/ Sbt 2035 | | | | | | | | | | | | | | | | | | |
| pT2037 | **pT2036** | I | G | T | S | I | **V** | | L | | P | I | | S | S | | L | T | | |
| pT2038 | **pT2032** | I | G | T | S | I | A | | I | | S | S | Q | S | S | K | L | T | | |
| pT2039 | **pT2032** | I | G | T | S | I | A | | I | | S | S | P | E | S | K | L | T | | |
| pT2040 | **pT2026** | I | G | T | S | I | **V** | | I | | P | I | | S | S | | L | T | | |
| pT2041 | **pT2035** | **L** | G | T | S | I | T | | L | | P | I | | S | S | | L | T | | |
| pT2042 | **pT2036** | **L** | G | T | S | I | T | | L | | P | I | | S | S | | L | T | | |
| pT2043 | **pT2040** | I | G | T | S | I | **V** | | I | | P | I | | S | S | | L | T | | |
| pT2044 | **pT2043** | **I** | G | T | S | I | V | | **V** | | P | I | | S | S | | L | T | | |
| pT2045 | **pT2043** | **I** | G | T | S | I | V | | **A** | | P | I | | S | S | | L | T | | |
| pT2046 | **pT2043** | **I** | G | T | S | I | V | | **G** | | P | I | | S | S | | L | T | | |
| pT2047 | **pT2043** | **L** | G | T | S | I | V | | I | | P | I | | S | S | | L | T | | |
| pT2048 | **pT2043** | I | G | T | S | I | V | | I | | **S** | I | | S | S | | L | T | | |
| pT2049 | **pT2044** | I | G | T | S | I | V | | V | | **S** | I | | S | S | | L | T | | |
| pT2050 | **pT2045** | I | G | T | S | I | V | | A | | **S** | I | | S | S | | L | T | | |
| pT2051 | **pT2001** | | S | | | | | | | | | | | | | | | | **G** | |
| pT2052 | **pT2004** | | A | | | | | | | | | | | | | | | | **G** | |

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| pT2053 | **pT2046** | **L** | G | T | S | I | V | | G | | P | I | | S | S | | L | T | | |
| pT2054 | **pT2050** | **I** | G | T | S | I | V | | A | | S | I | | S | S | K | L | T | | |
| pT2055 | **pT2048 with I11G in pro-region** | I | G | T | S | I | V | | I | | **S** | I | | S | S | | L | T | | |
| pT2056 | **pT2031** | I | G | T | S | I | A | | I | | S | S | | S | **D** | A | L | T | | |
| pT2057 | **pT2031** | I | G | T | S | I | **Y** | | **A** | | S | S | | S | S | A | L | T | | |
| | | | | | | | | | | | | | | | | | | | | | |
| | | 30 | 32 | 33 | 62 | 68 | 104 | 105 | 107 | 108 | 125 | 128 | 129 | 130 | 132 | 134 | 135 | 166 | 168 | 221 |
| pT2058 | **pT2031** | I | G | T | S | I | **Y** | | **G** | | S | S | | S | S | A | L | T | | |
| pT2059 | **pT2031** | I | G | T | S | I | **Y** | | **I** | | S | S | | S | S | A | L | T | | |
| pT2060 | **pN1013** | L | T | T | N | I | | | | | P | | | | | | | | | |
| pT2061 | **pN1014** | L | T | T | N | I | | | | | P | | | | | | | T | | |
| pT2062 | **pT2031** | I | G | T | S | I | A | | D | | S | S | | S | S | A | L | T | | |
| pT2063 | **pT2031** | I | G | T | S | I | A | | I | | S | S | | S | S | A | **S** | T | | |
| pT2064 | **pT2062** | I | G | T | S | I | A | | D | | S | S | | S | S | A | **S** | T | | |
| pT2065 | **pT2031** | I | G | T | S | I | A | | D | | S | S | | S | S | A | L | T | | |
| pT2066 | **pT2031** | I | G | T | S | I | A | | I | | S | S | | S | S | A | **S** | T | | |
| pT2067 | **pT2065** | I | G | T | S | I | A | | D | | S | S | | S | S | A | **S** | T | | |
| pT2068 | **pT2057 with S105C** | I | G | T | S | I | **Y** | **C** | **A** | | S | S | | S | S | A | L | T | | |
| pT2069 | **pT2058 with S105C** | I | G | T | S | I | **Y** | **C** | **G** | | S | S | | S | S | A | L | T | | |
| pT2070 | **pT2031 with S132F** | I | G | T | S | I | A | | I | | S | S | | S | **F** | A | L | T | | |
| pT2071 | **pT2057 with A134K** | I | G | T | S | I | **Y** | | **A** | | S | S | | S | S | K | L | T | | |
| pT2072 | **pT2059 with S105C** | I | G | T | S | I | **Y** | **C** | **I** | | S | S | | S | S | A | L | T | | |
| pT2073 | **pT2059 with I107V** | I | G | T | S | I | Y | | **V** | | S | S | | S | S | A | L | T | | |
| pT2074 | **pT2031 with FKAM cleave** | I | G | T | S | I | A | | I | | S | S | | S | S | A | L | T | | |
| pT2075 | **pT2068 with LKAM cleave** | I | G | T | S | I | Y | C | A | | S | S | | S | S | A | L | T | | |
| pT2076 | **pT2059 with A134C** | I | G | T | S | I | **Y** | | **I** | | S | S | | S | S | C | L | T | | |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| pT2077 | **pT2074 with I107V** | I | G | T | S | I | A | | **V** | | S | S | | S | S | A | L | T | | |
| pT2078 | **pT2073 with A134K** | I | G | T | S | I | Y | | **V** | | S | S | | S | S | K | L | T | | |
| pT2079 | **pT2059 with FKAM cleave** | I | G | T | S | I | Y | | I | | S | S | | S | S | A | L | T | | |
| pT2080 | **pT2079 with A134K** | I | G | T | S | I | Y | | I | | S | S | | S | S | **K** | L | T | | |
| pT2081 | **pT2073 with S105C** | I | G | T | S | I | Y | **C** | V | | S | S | | S | S | A | L | T | | |
| pT2082 | **pT2077 with S105C** | | | | | | | | | | | | | | | | | | | |
| pT2083 | **pT2070 with I107V** | I | G | T | S | I | A | | **V** | | S | S | | S | F | A | L | T | | |
| pT2084 | **pT2081 with FKAM cleave** | I | G | T | S | I | Y | C | V | | S | S | | S | S | A | L | T | | |
| pT2085 | **pT2083 with S62N** | I | G | T | **N** | I | A | | V | | S | S | | S | F | A | L | T | | |
| pT2086 | **pT2083 with A134K** | I | G | T | S | I | A | | **V** | | S | S | | S | F | **K** | L | T | | |
| pT2087 | **pT2083 with L135V** | I | G | T | S | I | A | | V | | S | S | | S | F | A | **V** | T | | |
| pT2088 | **pT2086 with L135V** | I | G | T | S | I | A | | **V** | | S | S | | S | F | **K** | V | T | | |
| pT2089 | **pT2064** | I | G | T | S | I | A | | D | | S | S | | S | S | A | **S** | T | | |
| pT2090 | **pT1011 with AKAMcys tail** | | | | | | | | | | | | | | | | | | | |
| pT2091 | **pT2090 minus pro** | | | | | | | | | | | | | | | | | | | |
| pT2092 | **pN1006 with AKAMcys tail** | L | G | T | S | I | | | | | P | | | | | A | | T | | |
| pT2093 | **pT2087 with G32A** | I | **A** | T | S | I | A | | V | | S | S | | S | F | A | V | T | | |
| pT2094 | **pS164 with AKAMcys tail** | V | A | S | N | V | A | S | I | | S | S | P | S | S | A | L | S | P | A |
| pT2095 | **pT2087 with S125P** | | | | | | | | | | P | | | | | | | | | |
| pT2096 | **pT2093 with S125P** | | | | | | | | | | P | | | | | | | | | |
| pT2097 | **pT2032 with ACCC tail** | | | | | | | | | | | | | | | | | | | |

| | | 30 | 32 | 33 | 62 | 68 | 104 | 105 | 107 | 108 | 125 | 128 | 129 | 130 | 132 | 134 | 135 | 166 | 168 | 221 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| pT2098 | pT2032 with Avi tag | | | | | | | | | | | | | | | | | | | |
| pT2099 | | | | | | | | | | | | | | | | | | | | |
| pT2100 | pT2093 with LKAM | I | A | T | S | I | A | | V | | S | S | | S | F | A | V | T | | |
| pT2101 | pT2093 with FKAM | I | A | T | S | I | A | | V | | S | S | | S | F | A | V | T | | |
| pT2102 | pT2094 with DAKAM | V | A | S | N | V | A | S | I | | S | S | P | S | S | A | L | S | P | A |
| pT2103 | pT2094 with DLKAM | V | A | S | N | V | A | S | I | | S | S | P | S | S | A | L | S | P | A |
| pT2104 | pT2087 with V135A | I | G | T | S | I | A | | V | | S | S | | S | F | A | A | T | | |
| pT2105 | pT2092 with HKAMcys tail | L | G | T | S | I | | | | | P | | | | | A | | T | | |
| pT2106 | pT2101 with ACCC tail | I | A | T | S | I | A | | V | | S | S | | S | F | A | V | T | | |
| pT2107 | pT2086 with ACCC tail | I | G | T | S | I | A | | V | | S | S | | S | F | K | L | T | | |
| pT2108 | pT2095 with FKAM cleave | I | G | T | S | I | A | | V | | P | S | | S | F | A | V | T | | |
| pT2109 | pT2096 with FKAM cleave | I | A | T | S | I | A | | V | | P | S | | S | F | A | V | T | | |
| pT2110 | pT2101 I30V A32S | **V** | **S** | T | S | I | A | | V | | S | S | | S | F | A | V | T | | |
| pT2111 | pT2110 S125P | V | S | T | S | I | A | | V | | **P** | S | | S | F | A | V | T | | |
| pT2112 | pT2101 with L217A | I | A | T | S | I | A | | V | | S | S | | S | F | A | V | T | | |
| pT2113 | pT2101 with L217G | I | A | T | S | I | A | | V | | S | S | | S | F | A | V | T | | |
| pT2114 | pT2022 L135E | I | G | T | S | I | T | | V | | P | S | | S | S | | **E** | T | | |
| pT2115 | pT2022 L135Q | I | G | T | S | I | T | | V | | P | S | | S | S | | **Q** | T | | |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| pT2116 | **pT2087 with F132S** | I | G | T | S | I | A | | V | | S | S | | S | S | A | **V** | T | | |
| | | 30 | 32 | 33 | 62 | 68 | 104 | 105 | 107 | 108 | 125 | 128 | 129 | 130 | 132 | 134 | 135 | 166 | 168 | 221 |
| pT2117 | **pT2057 with P52R and Q103C** | I | G | T | S | I | **Y** | | **A** | | S | S | | S | S | A | L | T | | |
| pT2118 | **pT2083 with P52R andQ103C** | I | G | T | S | I | A | | **V** | | S | S | | S | F | A | L | T | | |
| pT2119 | **His tag-pT2003** | | D | | | | | | | | | | | | | | | | | G |
| pT2120 | redesign of loop region in subtilisin | | | | | | | | | | | | | | | | | | | |
| pT2121 | pT2022 with A134Q, L135R | I | G | T | S | I | T | | V | | P | S | | S | S | Q | R | T | | |
| pT2122 | pT2022 with A134Q, L135K | I | G | T | S | I | T | | V | | P | S | | S | S | Q | K | T | | |
| pT2123 | pT2022 | I | G | T | S | I | **S** | | **I** | **T** | P | S | | S | S | **D** | **D** | T | | |
| pT2124 | pT2022 | I | G | T | S | I | **S** | | V | **T** | P | S | | S | S | **D** | **D** | T | | |